ML Reading Group:
The Algorithmic Foundations of Differential Privacy

March 2022

by Cynthia Dwork and Aaron Roth

Kevin Sheppard

March 8, 2022

Key Idea

- An algorithm is differentially private if the removal of a single observation as a negligible effect on the output.
  - The researcher must know no more about a single individual after the analysis than before.
- Aggregates are not generally private
  - The mean is not differentially private
    - n observations, $n-1$ of the individuals could collude to know the value of the missing observation
    - The mean of subsets reveal group averages

Lingo

Definition (Curator)

The curator holds the true data of individuals in database $\mathcal{D}$. Data for each individual is held in a single row.

Definition (Non-interactive Access)

The curator produces a transformation of $\mathcal{D}$ which is then available to researchers.

Definition (Query)

A query is a function applied to a database, $f(\mathcal{D})$.

Lingo

Definition (Online Access)

The researcher can submit queries to the database and can condition on the responses recieved when deciding whic query to run.

Definition (Privacy Mechanism)

An algorithm that takes an input database, a universe $\mathcal{X}$ is data types (all possible values for rows, whether in the data base or not), random noise, and produces an output string.

Classic Example

Answering a yes/no question

- Are you HIV positive?

Algorithm

Flip a coin

1. *If tails, answer truthfully*
2. *If heads, flip a second coin and answer yes if heads and no if tails*

- Researcher does not see any of the coin flips

Classic Example

Answering a yes/no question

- Allows researcher to learn about aggregate without knowing individual status
- If true rate is p, then

$$\lim_{n\to\infty} \frac{1}{n} \sum X_i = \frac{1}{4} + \frac{p}{2} \to \hat{p} = 2\bar{X} - \frac{1}{2}$$

- This estimator is less efficient than the MLE in the no privacy case
- Able to protect individual privacy except in edge cases where $p \in (0, 1)$
- While the individual status is protected, the researcher *does* learns something about the individual through inference on the group.
  - This learning can be bad for the individual, e.g., higher insurance premiums based on the analysis
  - The key point is that the researcher would learn this in a large sample whether the individual particiapted or not

Harder Stuff

Definition

Given a discrete set B, the probability simplex over B, $\Delta\left(B\right)$ is defined to be

$$\Delta\left(B\right) = \left\{ x \in \mathcal{R}^{|B|} : x \geq 0 \text{ for all } i \text{ and } \sum_{i=1}^{|B|} x_i = 1 \right\}$$

Definition

A randomized algorithm $\mathcal{M}$ with domain A and range B is associated with a mapping $M : A \to \Delta B$ such that $M\left(a\right) = b$ with probability $M\left(a\right)_b$ for each $b \in B$

Database Distance

Definition

The l_1 norm of a database x is

$$||x||_1 = \sum_{i=1}^{|X|} |x_i|$$

is the number of rows in the database. The l_1 distance between two database $||x - y||_1$ is the number of rows where the two databases differ.

Key Definition

Definition

A randomized algorithm $\mathcal{M}$ with domain $\mathbb{N}^{|\mathcal{X}|}$ is (ϵ, δ)-differentially private if for all $\mathcal{S} \subseteq$ Range $(\mathcal{M})$ and for all $x, y \in \mathbb{N}^{|\mathcal{X}|}$such that $||x - y||_1 = 1$if

$$Pr\left(\mathcal{M}\left(x\right) \in \mathcal{S}\right) \leq \exp\left(\epsilon\right) \Pr\left(M\left(y\right) \in \mathcal{S}\right) + \delta$$

- δ should be small, $< 1/||x||$
- 0 is simplest to think about

Privacy Loss

Definition

The privacy loss of observing some output ξ between two databases x and y is

$$\mathcal{L}^{\xi}_{\mathcal{M}(x)||\mathcal{M}(y)} = \ln \frac{\Pr\left(\mathcal{M}\left(x\right) = \xi\right)}{\Pr\left(\mathcal{M}\left(y\right) = \xi\right)}$$

- The privacy loss over all ξ in an (ϵ, δ)-differentially private algorithm is less than ϵ with probability $1 - \delta$.

Necessity of Random Noise

Laplace Mechanism

- Consider adding a draw from a Laplace (double exponential) distribution
- l_1 sensitivity of a function is

$$\Delta f = \max_{x,y:||x-y||_1=1} ||f(x) - f(y)||_1$$

- Laplace mechanism considers algorithm

$$\mathcal{M}(x, f, \epsilon) = f(x) + (Y_1, \ldots, Y_d)$$

where $Y_j \sim \text{Lap}(\Delta f / \epsilon)$

Common Moments

Mean

$$\Delta f = \frac{1}{n}$$

adding $Y \sim \text{Lap}\left(1/n\epsilon\right)$ to an average provides differential privacy.

Standard Deviation

$$\Delta f = \frac{1}{\sqrt{n}}$$

adding $Y \sim \text{Lap}\left(1/\sqrt{n}\epsilon\right)$ to a standard deviation provides differential privacy.

- **Notes**
  - Assume data to be appropriate normalized
  - Necessary for ϵ and δ to be meaningful

Regression

Based on Zhang, et. al. (2012, arXiv:1208.0219v1)

- Data is tuple $z_i = (y_i, x_{1i}, \ldots, x_{di})$
- Assumed to be normalized

$$\sqrt{\sum_{j=1}^{d} x_{ji}^2} < 1$$

- Interest in

$$\hat{\beta} = \operatorname{argmin} \sum_{i=1}^{n} \left(y_i - x_i'\beta\right)^2$$

- We know OLS is sensitive to adding noise to x: attenuation bias

Regression

- Appeal to Stone–Weierstrass Theorem to provide general solution
- Continuously differentially function can always be written as a (potentially infinite) polynomial of β

$$f\left(z_i, \beta\right) = \sum_{j=0}^{J} \sum_{\phi \in \Phi_j} \lambda_{\phi z_i} \phi\left(\beta\right)$$

where

$$\Phi_j = \left\{\beta_1^{c_1} \beta_2^{c_2} \ldots \beta_d^{c_d} : \sum c_i = j \text{ and } c_i \in \mathbb{N}\right\}$$

- Trivial for OLS since exact second order representation. Extends to logistic regression with truncation.

Key Idea

- Protect privacy by perturbing polynomial coefficient $\lambda_{\phi z_i}$
- The OLS objective is

$$\sum_{i=1}^{n} y_i^2 - \sum_{j=1}^{d} \beta_j \sum_{i=1}^{n} 2 y_i x_{ij} + \sum_{k=1}^{d} \sum_{l=1}^{d} \beta_k \beta_l \sum_{i=1}^{n} x_{ki} x_{li}$$

$$= \sum_{i=1}^{n} y_i^2 - \sum_{j=1}^{d} \beta_j \lambda_{\phi_{1j} z} + \sum_{k=1}^{d} \sum_{l=1}^{d} \beta_k \beta_l \lambda_{\phi_{2kl} z}$$

- In this model,

$$\Delta = 2 \max \sum_{j=1}^{2} \sum_{\phi \in \Phi_j} ||\lambda_{\phi z_i}||$$

$$\leq 2\left(1 + 2d + d^2\right) = 2\left(d+1\right)^2$$

- Follows directly from normalization assumption

Privacy Protection

Algorithm (Functional Mechanism)

1. *Set* $\Delta = 2\left(d+1\right)^2$
2. *For* $j \in \{0,1,2\}$
    a. *For* $\phi \in \Phi_j$ *set* $\lambda_\phi = \left(\sum_{i=1}^{n} \lambda_{\phi z_i}\right) + \textit{Lap}\left(\Delta/\epsilon\right)$
3. *Set objective to* $\sum_{j=1}^{2} \sum_{\phi \in \Phi_j} \lambda_\phi \phi\left(\beta\right)$
4. *Estimate parameters* $\bar{\beta}$ *by minimizing objective in 3*

Remarks

- When n is large objective is very close to OLS objective
- Consistent under standard setups despite noise
- Essentially iid perturbations of covariance matrix elements
  - Different from perturbations of x since can be either sign
  - Perturbation is $O_p\left({}^1/_n\right)$
- Protects ϵ-differential privacy
- No discussion of inference – satisfied with consistency
  - Privacy Protected statistics are probably nearly sufficient to implement homoskedastic inference
  - Only require sample size

Thoughts and Questions

- Protecting privacy in offline mode seems relatively easy for a fixed estimator
  - In practice requires curator to be available for researchers to handle general analysis
  - Can replace with a bot curator?
- What about different databases?
- It is practical to also general (unrestricted) analysis while preserving privacy?
- Differential privacy often limit themselves to standard estimators. Shouldn't the problem be a joint optimization problem across all privacy mechanisms and consistent (or other sensible) estimators?
- Limits the ability to apply improved estimators to existing databases since privacy protected sufficient statistics" might not be available.
- Are some classes of estimators easier to use while protecting privacy (e.g., rank-based statistics)?
- Given a privacy mechanism has been applied, how close are these standard statistics to the MLE?
- Privacy feel necessary but not sufficient for rich, non-NDA data sharing. Vendor reputation risk is still present.